03 - Multivariate Normal Distribution

Junvie Pailden

SIUE, F2017, Stat 589

August 30, 2017

Multivariate Normal Distribution

$\mathbf{X} = [X_1, \ldots, X_p]'$ has a p-dimensional normal distribution with $\mu = E(\mathbf{X})$ and $\Sigma = Var(\mathbf{X})$.

Density function

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$

The quantity $(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$ is called

- a squared Mahalanobis distance of $\mathbf{x}$ from μ
- a quadratic form
- statistical distance of $\mathbf{x}$ from μ

Notation: $\mathbf{X} \sim N_p(\mu, \Sigma)$

Multivariate Normal Distribution

The density function does not exist when

- Σ is not positive definite
- $|\Sigma| = 0$ (determinant is zero)
- Σ^{-1} does not exists (singular)

We assume that Σ is positive definite, i.e.

$$\mathbf{a}'\Sigma\mathbf{a} > 0$$

for every non-zero $p \times 1$ vector $\mathbf{a}$ of real numbers.

The MVN distribution belongs to the family of elliptical distributions. In two and three dimensional case, the joint distribution forms an ellipse and an ellipsoid.

MVN Computations in R using mvtnorm package

- dmvnorm to compute density function values
- pmvnorm to compute probability values
- rmvnorm to generate values

```
library(mvtnorm)
# density at (0,0) of standard bivariate MVN
dmvnorm(x = c(0,0))
```

```
# [1] 0.1591549
```

```
# density at (0,0) of bivariate MVN with mean (1,1)
# cov diag(2,2)
dmvnorm(x = c(0,0), mean=c(1,1), sigma = diag(2,2))
```

```
# [1] 0.04826618
```

Assume that $\mathbf{X} = [X_1, X_2, X_3]'$ is MVN with mean $\mu = [0, 0]'$ and covariance

$$\Sigma = \begin{bmatrix} 1 & 3/5 & 1/3 \\ 3/5 & 1 & 11/15 \\ 1/3 & 11/15 & 1 \end{bmatrix}$$

We are interested in the probability

$$\Pr(-\infty < X_1 \leq 1, -\infty < X_2 \leq 4, -\infty < X_3 \leq 2)$$

```
sigma1 <- matrix(c(1, 3/5, 1/3, 3/5, 1,
                   11/15, 1/3, 11/15, 1), nrow = 3)
pmvnorm(mean = c(0, 0, 0), sigma = sigma1,
        lower = c(-Inf, -Inf, -Inf), upper = c(1, 4, 2))
```

```
# [1] 0.8279846
# attr(,"error")
# [1] 4.349239e-07
# attr(,"msg")
# [1] "Normal Completion"
```

Geometry for the Bivariate Normal Distribution

- Consider $\mathbf{X} = [X_1, X_2]' \sim N_p(\mu, \Sigma)$, where $\mu = [\mu_1, \mu_2]'$ and

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_{22} \end{bmatrix}$$

where $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ and $\sigma_1 = \sqrt{\sigma_{11}}$, $\sigma_2 = \sqrt{\sigma_{22}}$.

- We can write

$$|\Sigma| = \sigma_{11}\sigma_{22}(1-\rho^2), \quad \Sigma^{-1} = \frac{1}{|\Sigma|}\begin{bmatrix} \sigma_{22} & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_{11} \end{bmatrix}$$

Bivariate Normal Density Function

$$\phi(x_1,x_2) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

- This density is well defined if $-1 < \rho < 1$.
- If $\rho = 0$, then $\phi(x_1, x_2) = \phi(x_1) \cdot \phi(x_2)$ where $\phi(x_i|\mu_i, \sigma_i)$ is the pdf of univariate normal with mean μ_i and standard deviation σ_i.

Uncorrelated $\iff$ *independence (only for multivariate normal)*

Bivariate Normal Density Function

The density is constant for $\mathbf{x} = [x_1, x_2]'$ points for which (c is constant)

$$c = \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right]$$

- This is an equation for an ellipse centered at $\mu = [\mu_1, \mu_2]'$.
- What are the lengths and positions of the major axes of the ellipsoids corresponding to contours of constant density $(\rho \neq 0)$?

Bivariate Normal, Eigenvalues

Eigenvalues of Σ for Bivariate Normal (when $\sigma_{11} = \sigma_{22}$) are the solutions to

$$
\begin{aligned}
0 = |\Sigma - \lambda I| &= \begin{vmatrix} \sigma_{11} - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11} - \lambda \end{vmatrix} \\
&= (\sigma_{11} - \lambda)^2 - \sigma_{12}^2 \\
&= (\sigma_{11} - \lambda - \sigma_{12})(\sigma_{11} - \lambda + \sigma_{12})
\end{aligned}
$$

The eigenvalues are

$$
\begin{aligned}
\lambda_1 &= \sigma_{11} + \sigma_{12} \\
\lambda_2 &= \sigma_{11} - \sigma_{12}
\end{aligned}
$$

The first eigenvalue-eigenvector pair is

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \;\; e_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

The second eigenvalue-eigenvector pair is

$$\lambda_2 = \sigma_{11} - \sigma_{12}, \;\; e_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

When $\sigma_{12} > 0$, λ_1 is the largest eigenvalue, and its associated eigenvector lies in the 45 deg line through the $\mu = [\mu_1, \mu_2]'$.

The ratio of the lengths of the axes is

$$\frac{\text{length of the major axis}}{\text{length of the minor axis}} = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}$$

Example: Bivariate Normal Distribution

Consier the Bivariate Normal Distribution with

```
mu0 <- c(0, 0) # population mean
mu0
```

```
# [1] 0 0
```

```
# pop'n covariance matrix
# sigma11 = sigma22 = 1, rho = sigma12 = 0.79
Sigma0 <- matrix(c(1, .79, .79, 1), 2)
Sigma0
```

```
#      [,1] [,2]
# [1,] 1.00 0.79
# [2,] 0.79 1.00
```

Eigenvalues and eigenvectors of the covariance matrix

```
e <- eigen(Sigma0) # compute eigenvalues/eigenvectors
lambda <- e$values  # eigenvalues only
lambda
```

```
# [1] 1.79 0.21
```

```
evec <- e$vector  # eigenvectors only
evec
```

```
#           [,1]       [,2]
# [1,] 0.7071068 -0.7071068
# [2,] 0.7071068  0.7071068
```

Since $\sigma_{11} = \sigma_{22} = 1$ with $\sigma_{12} = 0.79 > 0$ and $\lambda_1 = 1.79$ is largest, then the first eigenvector lies in the 45 deg line through the mean $(0, 0)$.

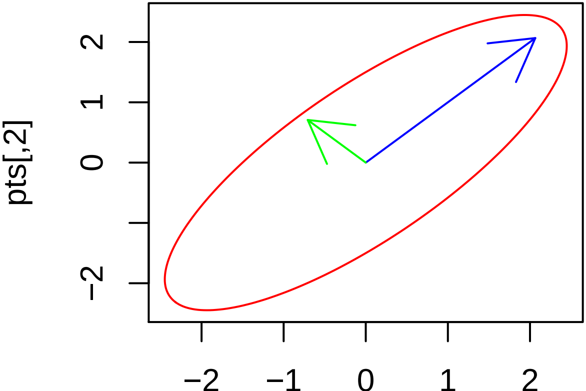
The length of the major axis is proportional to the root of the largest eigenvalue.

```
sqrt(lambda[1])/sqrt(lambda[2])
```

```
# [1] 2.919556
```

Major-axis is close to thrice as long as the minor-axis.

```r
# ellipse function is in mixtools package
mixtools::ellipse(mu0, Sigma0, newplot = TRUE,
            type = "l", col = "red") # curve, color
# first eigenvector, major-axis is 2.92 times longer
arrows(0, 0, 2.92*evec[1,1], 2.92*evec[2,1], col = "blue")
# second eigenvector
arrows(0, 0, evec[1,2], evec[2,2], col = "green")
```



Generating Multivariate Normal Samples

Use `mvtnorm::rmvnorm` generate mult. normal sample points.

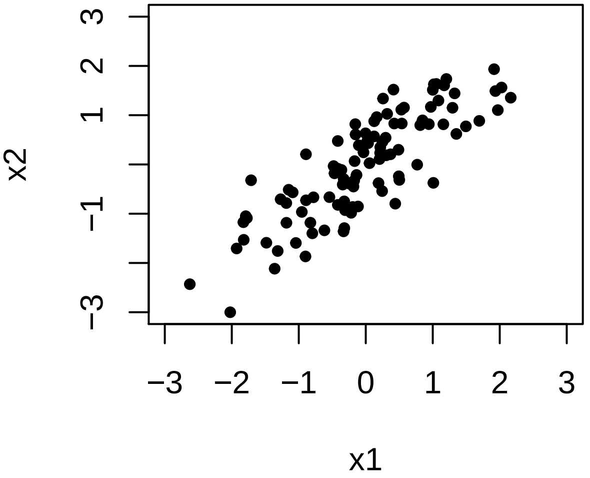
```r
set.seed(21) # set a seed # to get the same points
# generate 100 points using `mvnorm`
X.samp <- rmvnorm(100, mean = mu0, sigma = Sigma0)
colnames(X.samp) <- c("x1", "x2") # change colnames
colMeans(X.samp) # sample means
```

```
#          x1          x2
# -0.02177390 -0.01724441
```

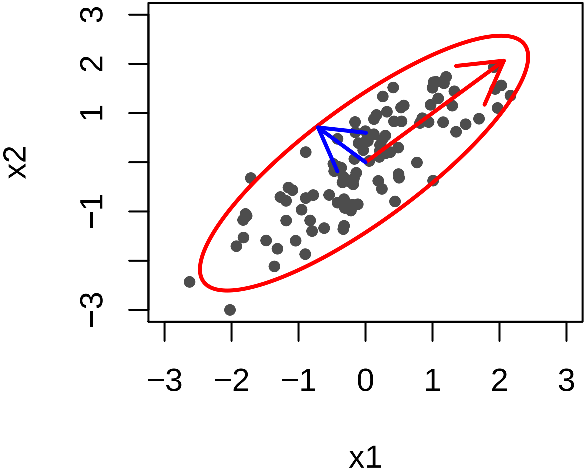
```r
cov(X.samp)   # sample covariance
```

```
#           x1        x2
# x1 1.0015610 0.8827785
# x2 0.8827785 1.1190363
```

```r
plot(X.samp, # sample data
     pch = 20, # solid dot points
     xlim = c(-3, 3), ylim = c(-3, 3)) # set axis limits
```



Add the eigenvectors and 95% ellipse band



Kernel Density Estimation (KDE)

- Probability histograms are density estimates in the sense that it approximates the shape of true density of the data.
- KDE allows us to estimate (using kernels or small density functions) the density from which each sample was drawn.
- Check this link for more information.
- We use the `kde2d()` function in the `MASS` package to construct KDE's for bivariate distributions.

Apply the KDE

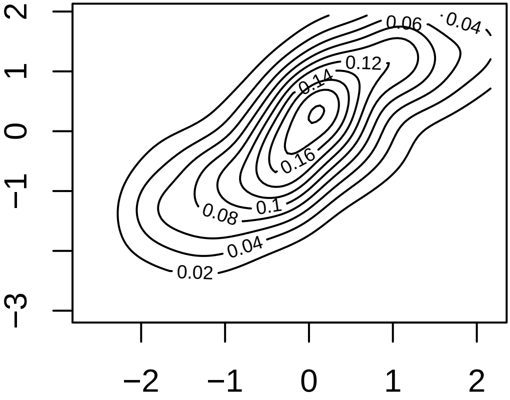
```r
# we obtain a kernel density estimate of X.samp
X.kde <- MASS::kde2d(X.samp[,1], X.samp[,2], n = 100)
str(X.kde) # check structure of kde output
```

```
# List of 3
#  $ x: num [1:100] -2.63 -2.58 -2.53 -2.48 -2.43 ...
#  $ y: num [1:100] -3 -2.95 -2.9 -2.85 -2.8 ...
#  $ z: num [1:100, 1:100] 0.00685 0.00741 0.00797 0.00853
```

- The points (x, y) forms the grid of points over the data support.
- The value z is the estimate of the bivariate normal density $\phi(x, y)$ at specific points (x, y).

Plot the contours

```
# plot the contours of the kde output
contour(X.kde)
```



Fancier 2d Visualization

```
image(X.kde) # use image function to create a base plot
contour(X.kde, add = T) # add the contours
```